COMMENTARY

Can the EU become a Global Norm Setter in Ethical AI?

VOCAL
EUROPE

*This commentary was written by **Matthias Peschke** | 25 November 2019

Rue de la Science 14, 1040 Brussels

office@vocaleurope.eu

+ 32 02 588 00 14

VOCAL EUROPE

RUE DE LA SCIENCE 14B, 1040 BRUSSELS
TEL: +32 02 588 00 14
VOCALEUROPE.EU

TWITTER.COM/THEVOCALEUROPE
FACEBOOK.COM/VOCALEUROPE
YOUTUBE.COM/VOCALEUROPE
INSTAGRAM.COM/VOCALEUROPE

Disclaimer and Copyright

This document is prepared for, and addressed to Vocal Europe and its audience. The content of the document is the sole responsibility of its author(s) and any opinions expressed herein should not be taken to represent an official position of Vocal Europe. Reproduction and translation for non-commercial purposes are authorised, provided the source is acknowledged.

Summary

There is a growing debate about whether AI should be regulated on the basis of ethical principles. The EU with its ethical guidelines has taken up a progressive step in this respect and received mixed reviews with sceptics fearing that regulation will end up limiting AI technology and harm society by depriving it of the overwhelmingly positive potential AI provides. Proponents of regulation, by contrast, focus on privacy concerns and unintended consequences that have already taken a toll especially among the most vulnerable in society by reinforcing existing human biases.

The risks of AI can be divided up into long and short-term risk, an important distinction that many reports fail to understand. Long-term risk, also known as existential risk, revolves around the question of what will happen to society when the intellect of a computer supersedes those of human beings while short-term risks address the current challenges of data-driven algorithms including biases, discrimination and privacy violations.

As a result, ethical concerns have gradually increased and public institutions as well as the private sector look for ways to curb the negative impacts of this technology by setting up principles for human-centric AI development. However, since the black box model is still dominating the AI landscape, companies currently lack the capacity to implement these principles. Thus, further research into explainable AI and ethics are of vital importance in order to guarantee the responsible development of AI technology.

Can the EU become a Global Norm Setter in Ethical AI?

The debate about whether ethical considerations are obstructing innovative technology is increasing. The EU was the first to regulate big data and even exported this regulation since foreign companies had to update their regulation in order to maintain access to the European market. The Commission has followed up with ethical guidelines for trustworthy AI[1] reinforcing its desire to become a global norm setter in this field.

These efforts have received mixed reviews with critics pointing out that the EU will not be able to compete internationally by "philosophizing on the sidelines"[2]. Both sides make reasonable arguments which raises the question whether ethical considerations should already be prioritised by the EU.

Therefore, this commentary looks at some of the long and short-term challenges AI poses to society and illustrates how these problems are being addressed by global actors. The goal is to provide a general overview of the ethical debate and assess whether the EU can play a central role in shaping it.

Is AI a threat to civilisation?

The question of whether AI might be a threat to human society is a central point of conflict and often the cause for heated debate. One side dismisses this notion as nonsense because the current state of AI is a far cry from anything that could be a sincere threat to the survival of the human race. The scope of application will remain narrow for the foreseeable future and, thus, diffusing such a narrative is a form of alarmism that will ultimately harm the development of AI technology whose overwhelming potential is positive.[3]

It also distracts from the more imminent problems current AI applications cause regarding mass-surveillance and biased decision-making. A common target of this criticism is tech entrepreneur Elon Musk who repeatedly stated that AI is more dangerous than nuclear weapons and needs to be regulated by the state.[4]

Although this criticism has a substantiated basis, it often omits the underlying issue of what precisely is meant by existential risk from Artificial General Intelligence (AGI). This field has been the focus of research for decades and analyses the probable effects of an AI whose intellect supersedes those of humans. In 2014, one of the leading experts in this field, Nick Boström, published the seminal work "Superintelligence" which concludes that "once unfriendly superintelligence exists, it would prevent us from replacing it or changing its preferences. Our fate would be sealed."[5]

The book attracted lots of attention especially among the global AI community and mobilised investments into AI safety while, at the same time, starting a broader debate on the ethics of AI systems. Musk's doomsday scenario is, therefore, based on research done in the field of existential risk and it convinced him to found the non-profit organisation OpenAI with the goal of researching

[1] European Commission (2019): Ethics Guidelines for Trustworthy AI. Document by the High-Level Expert Group on AI. Document accessible here.

[2] Chivot, Eline/Castro, Daniel (2019): The EU's "softball" approach to Artificial Intelligence will lose to China's "hardball" | View. In: *Euronews.com* on 5 February 2019. Article accessible here.

[3] Clifford, Catherine (2018): Google billionaire Eric Schmidt: Elon Musk is 'exactly wrong' about A.I. because he 'doesn't understand'. In: *CNBC* on 29 May 2018. Article accessible here.

[4] Jeffries, Daniel (2017): Why Elon Musk is Wrong about AI. In: *Hackernoon* on 14 August 2017. Article accessible here.

[5] Boström, Nick (2014): Superintelligence. Paths, Dangers, Strategies. Oxford: Oxford University Press.

and democratising AI as well as counterbalancing the growing concentration of this technology at a couple of tech giants.[6] So far, predictions about when computers will exceed human intelligence vary a lot with the majority of researchers assuming that it will not happen before the end of this century.[7] On the one hand, this can be seen as confirmation that too much regulatory interference as of now would not be required.

However, on the other hand, it can also be argued that establishing ethical principles at this early stage is crucial to set out the right foundation. In the past, governments have failed to protect workers and the environment from the impact of large-scale industrialisation with devastating effects for the welfare of human beings and the planet itself.[8] The desire to learn from these historical negligences can, thus, also be seen as an important maturation of human society.

Big Data: Protecting privacy versus training algorithms

Whereas existential risk for humanity might be a long-term challenge, there are many topical issues that are highly-contentious such as, most notably, big data. The reason is simple: AI becomes better at predicting outcomes and recognising patterns the more data is fed to it. As a matter of fact, this process often clashes with the fundamental rights of EU citizens. Tesla, for example, uses the data created by its customers to improve its self-driving algorithm including location, camera footage and pedal-usage.[9]

Without access to such data, the technology would not be able to advance which makes more concerns about privacy result in slower adoption of innovative technology. In fact, it is argued that the country most likely to lead in autonomous vehicles will be the one with the most conducive legislation, not the one that leads the technological development.[10]

Thus, the business side is fearful of rigorous regulation as this might lead to a partial ban of the technology rather than a reduction of the negative impacts some AI applications might entail. Instead, it advocates for a data-driven approach concurring that a focus on data protection will be an obstacle for applying AI[11] which is already illustrated by the slow adoption of AI use cases among companies.[12]

Unsurprisingly, the EU's Data protection rules are seen in a negative light because the collection and repurposing of data are subject to strict rules. Companies whose production process heavily relies on data will be adversely affected in terms of economic output and productivity.[13] Criticism is also evoked by the requirement to explain how the gathered data is being used saying that, since algorithms process huge amounts of data in a black box, it is not possible to comprehend every suggestion and

[6] Piper, Kelsey (2018): Why Elon Musk fears artificial intelligence. In: *Vox* on 2 November 2018. Article accessible here.

[7] Vincent, James (2018): This is when AI's top researchers think artificial general intelligence will be achieved. In: *The Verge* on 27 November 2018. Article accessible here.

[8] Taddeo, Mariarosaria/Floridi, Luciano (2018): How AI can be a force for good. An ethical framework will help to harness the potential of AI while keeping humans in control. In: *Science*, 361 (6404), 751-752.

[9] Muller, Johann (2019): What Tesla knows about you. In: *Axios* on 13 March 2019. Article accessible here.

[10] Lewis, Dev (2019): China's Techno-Utilitarian Experiments with Artificial Intelligence. Singapore: Konrad Adenauer Stiftung. Article accessible here.

[11] Burkert, Andreas (2017): Ethics and the Dangers of Artificial Intelligence. In: *ATZ worldwide*, 11 (2017), 8-13.

[12] Csernatoni, Raluca (2019): An Ambitious Agenda or Big Words? Developing a European Approach to AI. In: Egmont Policy Brief No. 117. Document accessible here.

[13] Ferracane, Martina Francesca et al. (2018): Do Data Policy Restrictions Impact the Productivity Performance of Firms and Industries? ECIPE/DTE Working Paper 01. Document accessible here.

decision they come up with. Favouring explainability would reduce accuracy of AI systems and go to the detriment of businesses that would be unable to provide cutting-edge services to customers.[14] Furthermore, the argument goes that decisions made by humans are also not always completely explainable and holding AI to higher standards than humans would be irrational and continue discouraging innovation. The core benefit of AI is to allow for greater efficiencies but if humans are to review every step on the way, then these efficiencies can hardly materialise.[15]

It, therefore, comes as no surprise that the leading countries in the development and application of AI technology are those with fewer data protection rules. Accordingly, PwC estimates that the economy in 2030 will be 14 percent higher as a result of AI with the greatest gains accruing in China and North America.[16]

There is no doubt that missing out on these developments would make states less influential on the international stage. However, for liberal democracies, it is equally important to look at the negative implications for liberal rights. Whereas having reservations is dismissed by some as technophobia or "speculative concerns",[17] "a big part of ethics," as explained by Rachel Thomas, Director of the USF Center for Applied Data Ethics, "is to think about what could go wrong before it does."[18]

Moreover, there already exist many examples of algorithms that have led to harmful outcomes and unintended consequences. Cathy O'Neil has shed light on some of these issues as early as 2016 when she published her book "Weapons of Math Destruction".[19] She demonstrated that data can be misused in a variety of ways such as in teacher evaluations, shift scheduling and predatory advertisement. Each of these applications want to optimise processes but, in doing so, they often prey on the most vulnerable in society. Techniques like these have even found their way into political campaigns with the most recent Cambridge Analytica Scandal just being the tip of the iceberg.[20]

Some algorithms, although launched with good intentions, end up contributing to a vicious feedback loop that systematically reinforces existing inequalities instead of disarming them. Examples for that are often found in the field of law enforcement where the use of predictive policing has often simply targeted minorities and the poor more intensely rather than effectively created safer communities.

The reason for this is that, in general, algorithms struggle with understanding concepts. It is difficult to quantify fairness, justice or equality which begs the question "whether we as a society are willing to sacrifice a bit of efficiency in the interest of fairness."[21]

An important aspect of fairness and an inherent principle of liberal democracies is holding decision makers to an account. Companies, by contrast, have to focus on profits which are increasingly generated by ever more sophisticated algorithms. For these companies, more transparency could expose the inner workings of revenue-driving algorithms and constitute a danger to a highly-lucrative business model.

[14] See footnote 11.

[15] Castro, Daniel/McLaughlin, Michael (2019): Ten Ways the Precautionary Principle Undermines Progress in Artificial Intelligence. In: *ITIF* on 4 February 2019. Article accessible here.

[16] PwC (2017): Sizing the prize: What's the real value of AI for your business and how can you capitalise? Report accessible here.

[17] See footnote 15.

[18] TWIMLcon (2019): Operationalizing responsible AI. Panel discussion. Podcast accessible here.

[19] O'Neil, Cathy (2016): Weapons of Math Destruction. How Big Data increases Inequality and threatens Democracy. New York: Crown.

[20] Wetsby, Joe (2019): 'The Great Hack': Cambridge Analytica is just the tip of the iceberg. In: *Amnesty Tech* on 24 July 2019. Article accessible here.

[21] See footnote 19.

Hence, there is a clear point of contention between the ideal outcome for open societies and for businesses. Delegating decisions to algorithms can lead to increased efficiencies but it would also blur the chain of responsibility and result in an erosion of public trust.[22] These external costs to society need to be taken into consideration as they will remain unaddressed if no regulation is in place.

Similar to pollution caused by factories, we are dealing with a market failure because the polluter does not have to bear the true costs of his actions. Rectifying this shortcoming is not an easy task not only because the true costs to society are difficult to assess but also because explainable artificial intelligence (XAI) is indeed a technological challenge that is still in its infancy.[23]

This brings us to the political component that is often forgotten when less regulation is demanded. Things that might make sense from an economical perspective, such as austerity and a balanced national budget, do not always yield the desirable social outcome. An example for this are the austerity measures implemented between 2010 and 2015 in the UK which have paved the way for the Brexit vote.[24]

Moreover, EU member states are already feeling a substantial populist backlash resulting from a variety of unaddressed cultural and economic grievances. Ignoring the privacy and ethical concerns of the population with respect to the digital transformation and AI can, thus, be considered ill-advised. Instead, ethical concerns should be addressed under the participation of relevant stakeholders in order to ensure that regulation does not stifle innovation but rather stirs it into a direction that is compatible with fundamental rights and values.[25]

The Ethical EU: A lone warrior?

While the fear of many businesses is that the EU Ethics Guidelines might lead to overregulation and discourage innovation, it is worth mentioning that the basic idea of having a human-centric approach to AI is generally welcomed by the majority of stakeholders. A recent study found that 94% of decision-makers in the IT sector agree that more attention should be paid toward developing responsible AI systems while 87% say that AI should be regulated.[26]

Businesses like Microsoft and Google have even come forward with their own ethical principles which seek to maximise AI's benefit for humanity and aim to respect fundamental values of fairness, transparency, privacy and accountability while also promising to not pursue potentially harmful uses.[27] [28] Even Facebook, though certainly pressured due to its involvement in the Cambridge Analytica Scandal, is funding ethical research into AI at the Technical University of Munich.[29]

[22] Open Society Foundation (2019): A Human-Centric Digital Manifesto for Europe. How the Digital Transformation can serve the Public Interest. Brussels: Open Society European Policy Institute/The European Consumer Organisation.

[23] Barredo Arrieta et al. (2019): Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI. In: *arXiv*: 1910.10045 [cs.AI]. Paper available here.

[24] Fetzer, Thiemo (2019): Did Austerity cause Brexit? CESifo Working Paper Series 7159. Munich: CESifo Group.

[25] More than 500 companies and organisations were asked to comment on the draft of EU Ethics Guidelines for Trustworthy AI before the final version was published.

[26] Snaplogic (2019): The AI Ethics Deficit — 94% of IT Leaders Call for More Attention to Responsible and Ethical AI Development. Study conducted by Vanson Bourne. Article available here.

[27] Microsoft (2019): Microsoft AI Principles. Text accessible here.

[28] Google (2019): Artificial Intelligence at Google: Our Principles. Text accessible here.

[29] Candela, Joaquin Quiñonero (2019): Facebook and the Technical University of Munich Announce New Independent TUM Institute for Ethics in Artificial Intelligence. In: *Facebook Newsroom* on 20 January 2019. Article accessible here.

By the same token, the World Economic Forum (WEF), which is comprised of more than a thousand multinational corporations and usually advocates in favour of deregulation, has also recognised the need to develop ethical standards to guarantee the benevolence of AI technology. Their recently published white paper is encouraging policymakers to take action with a variety of tools such as regulation, standardisation and certification.[30] The WEF even goes a step further by recommending states to emphasise ethics and data protection when drafting national strategies for AI.[31]

While this surely represents a step into the right direction, it is still unclear how strong the commitment to these non-binding principles is and whether these measures are undertaken to avoid hard-rule making by the state. Nevertheless, it shows that the ethical debate has gained traction and is unlikely to disappear anytime soon.

On the other side, international organisations as well as states have also taken the initiative and addressed ethical issues. Since 2017, the UN organises the Global Summit on AI for Good where all kinds of stakeholders come together to have a dialogue on how to harvest the benefits of AI to help realise the sustainable development goals. A substantial part of this discussion revolves around the question of how to guarantee that AI technologies are developed in a way that is trustworthy, safe and inclusive.

Likewise, the OECD is promoting ethical principles that seek to establish an international standard in the development of AI. Adopted by a total of 42 states, these principles resemble the EU's stance by calling for human-centric AI that is, among other things, trustworthy, explainable, transparent and accountable.[32]

Earlier this year, China has also come forward with the Beijing principles which acknowledge the necessity to develop a regulatory framework for AI that promotes the use of the technology for the good of humanity.[33] Its national AI strategy already emphasises the importance to strengthen research on legal, ethical and social issues related to AI and create systems that are traceable and accountable.[34] Moreover, President Xi recently elaborated that these efforts can only be fruitful under increased cooperation with all countries further underlining China's interest to be part of this global debate.[35]

Although this international consensus on the need to apply ethics is a step into the right direction, having non-binding mechanisms in place will not solve the problem on its own. Furthermore, as of now, companies and states do not have the experience and technical know-how to implement AI principles and avoid unintended negative consequences.[36]

Thus, setting up a legally-binding regulatory framework for AI needs to be accompanied by more research into explainable artificial intelligence. In combination, these efforts have the potential to ensure that the fundamental rights of individuals are protected and encourage innovation that truly benefits society.

[30] World Economic Forum (2019): AI Governance. A holistic approach to implement ethics into AI. Document accessible here.

[31] World Economic Forum (2019): A Framework for Developing a National Artificial Intelligence Strategy Centre for Fourth Industrial Revolution. Document accessible here.

[32] OECD (2019): Recommendation of the Council on Artificial Intelligence. Document accessible here.

[33] BAAI (2019): Beijing AI Principles. Text accessible here.

[34] Library of Congress (2019): Regulation of Artificial Intelligence: East/South Asia and the Pacific. Text available here.

[35] See footnote 10.

[36] See footnote 23.

Conclusion

Given the global desire to develop responsible AI systems, the EU's approach to ethical AI is reasonable and its ethical guidelines are set to influence and contribute to the global discussion of this issue. Being at the forefront of data protection, the EU is likely to take a leading role in the development of ethical principles as well.

Although the private sector has fears that ethical concerns will lead to overregulation and discourage innovation, the majority of stakeholders is welcoming the initiative to establish international standards. Moreover, countries across the globe are adopting ethical guidelines indicating an increasingly global consensus that states need to play a more active role in keeping the negative effects of this technology in check.

Even though it is unlikely that intelligent machines will supersede human intelligence in the coming decades, setting the right foundation at an early stage of development would be a wise decision. Yet, we need to recognise that AI has an overwhelming potential to contribute to human welfare which makes it so important that a regulatory framework curbs the negative effects of AI, not the technology itself.

For this to succeed, a global approach is necessary giving countries an opportunity to set aside differences and work together toward the common goal of establishing an international standard that reflects human values and enables human-centric innovation. If underpinned by increased research in the area of explainable artificial intelligence in order to tackle the underlying technological challenges, the long-term prospects for success are favourable.

In this context, we should keep in mind that the future is not set in stone but rather shaped by the paths we choose today. If we had conducted more research into climate change at an earlier stage and educated the population about the damaging potential, we probably wouldn't still have to debate the virtue of reducing CO_2 emissions.

Fact is, the ability to interpret large amounts of behavioural data is increasing the power of states and corporations over the individual. There are many ways in which algorithms already have negative impacts on people. Therefore, setting up sensible regulation, expanding research into ethics and explainable AI as well as educating the population on the benefits and dangers of the digital transformation is in the interest of everybody.

VOCAL EUROPE

RUE DE LA SCIENCE 14B, 1040 BRUSSELS
TEL: +32 02 588 00 14
VOCALEUROPE.EU

TWITTER.COM/THEVOCALEUROPE
FACEBOOK.COM/VOCALEUROPE
YOUTUBE.COM/VOCALEUROPE
INSTAGRAM.COM/VOCALEUROPE

Disclaimer and Copyright

This document is prepared for, and addressed to Vocal Europe and its audience. The content of the document is the sole responsibility of its author(s) and any opinions expressed herein should not be taken to represent an official position of Vocal Europe. Reproduction and translation for non-commercial purposes are authorised, provided the source is acknowledged.